

GenCore version 5.1.9
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2007, 15:06:57 ; Search time 182 Seconds
(without alignments)
381.770 Million cell updates/sec

Title: US-10-777-578-2

Perfect score: 770

Sequence: 1 MVVFTLLVLLVLLLTALC.....LQRLPQNSALLACRCAGAY 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA_Main.*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	770	100.0	150	3	US-09-895-836-2
2	770	100.0	150	3	US-09-893-737-230
3	770	100.0	150	4	US-10-777-578-2
4	770	100.0	150	5	US-10-970-713-230
5	164.5	21.4	246	5	US-10-450-763-59832
6	134	17.4	635	5	US-10-450-763-49851
7	134	17.4	2134	5	US-10-450-763-57983
8	132.5	17.2	138	5	US-10-450-763-36160
9	95.5	12.4	87	5	US-10-450-763-54124
10	93	12.1	87	5	US-10-450-763-43178
11	84	10.9	263	5	US-10-450-763-38317
12	84	10.9	263	5	US-10-450-763-54486
13	80	10.4	401	4	US-10-424-599-161124
14	77.5	10.1	105	4	US-10-437-963-113295
15	77.5	10.1	322	5	US-10-505-818-26
16	75.5	9.8	391	4	US-10-282-122A-63361
17	75	9.7	527	4	US-10-369-493-9634
18	75	9.7	947	4	US-10-437-963-130973
19	74.5	9.7	1153	4	US-10-437-963-108554
20	74.5	9.7	1153	4	US-10-732-923-15016
21	74	9.6	307	5	US-10-455-772-604
22	74	9.6	1128	4	US-10-425-115-368789
23	74	9.6	1442	4	US-10-437-963-200129
24	74	9.6	1466	4	US-10-437-963-200133
25	73.5	9.5	414	5	US-10-349-528-14
26	73	9.5	247	4	US-10-200-344-6
27	73	9.5	247	5	US-10-872-968-6

```

28 73 9.5 315 3 US-09-888-615-60 Sequence 60, Appl
29 73 9.5 315 6 US-11-037-243-60 Sequence 60, Appl
30 73 9.5 350 4 US-10-200-344-12 Sequence 12, Appl
31 73 9.5 350 5 US-10-872-968-12 Sequence 12, Appl
32 73 9.5 434 4 US-10-477-515-2 Sequence 2, Appl
33 73 9.5 437 4 US-10-200-344-10 Sequence 10, Appl
34 73 9.5 437 4 US-10-274-639-12 Sequence 12, Appl
35 73 9.5 437 4 US-10-333-574-12 Sequence 12, Appl
36 73 9.5 437 4 US-10-757-262-128 Sequence 128, Appl
37 73 9.5 437 5 US-10-872-968-10 Sequence 10, Appl
38 73 9.5 1225 5 US-10-741-849-7114 Sequence 7114, Ap
39 72.5 9.4 84 4 US-10-424-599-185950 Sequence 185950,
40 72.5 9.4 368 5 US-10-450-763-40386 Sequence 40386, A
41 72 9.4 296 4 US-10-025-806-88 Sequence 88, Appl
42 72 9.4 344 4 US-10-292-798-242 Sequence 242, App
43 72 9.4 344 4 US-10-297-021-8 Sequence 8, Appli
44 72 9.4 389 4 US-10-424-599-168332 Sequence 168332,
45 72 9.4 412 4 US-10-343-650A-378 Sequence 378, App

```

ALIGNMENTS

RESULT 1
US-09-895-836-2
; Sequence 2, Application US/09895836
; Publication No. US20020086367A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: NOVEL SECRETED PROTEINS
; FILE REFERENCE: 01-26
; CURRENT APPLICATION NUMBER: US/09/895, 836
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/215,446
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-836-2

Query Match 100.0%; Score 770; DB 3; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.7e-77;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVVFTLLVLLVLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQYFLPTGG 60
DB 1 MVVFTLLVLLVLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQYFLPTGG 60
QY 61 KCLHLRLTGQAFCLILEFLPWCNGIIESVLGLENRKVLGGSSQQMGARRGMWEVFP 120
DB 61 KCLHLRLTGQAFCLILEFLPWCNGIIESVLGLENRKVLGGSSQQMGARRGMWEVFP 120
QY 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150
DB 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150

RESULT 2
US-09-893-737-230
; Sequence 230, Application US/098993737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28

```
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 230
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-230

Query Match      100.0%; Score 770; DB 3; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.7e-77;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDKIKIDIQYFLPTGG 60
   |||||
Db 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDKIKIDIQYFLPTGG 60
   |||||

QY 61 KCLHLRLTGORAFCLIEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||
Db 61 KCLHLRLTGORAFCLIEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||

QY 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||
Db 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||

RESULT 3
US-10-777-578-2
; Sequence 2, Application US/10777578
; Publication No. US20040147722A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; APPLICANT: Taft, David W.
; TITLE OF INVENTION: NOVEL SECRETED PROTEINS
; FILE REFERENCE: 01-26
; CURRENT APPLICATION NUMBER: US/10/777,578
; PRIOR FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: US/09/895,836
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-777-578-2

Query Match      100.0%; Score 770; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.7e-77;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDKIKIDIQYFLPTGG 60
   |||||
Db 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDKIKIDIQYFLPTGG 60
   |||||

QY 61 KCLHLRLTGORAFCLIEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||
Db 61 KCLHLRLTGORAFCLIEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||

QY 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||
Db 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||

RESULT 4
US-10-970-713-230
; Sequence 230, Application US/10970713
; Publication No. US20050214791A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/10/970,713
; CURRENT FILING DATE: 2004-10-20
; PRIOR APPLICATION NUMBER: US/09/893,737
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 230
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-970-713-230

Query Match      100.0%; Score 770; DB 5; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.7e-77;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDKIKIDIQYFLPTGG 60
   |||||
Db 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDKIKIDIQYFLPTGG 60
   |||||

QY 61 KCLHLRLTGORAFCLIEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||
Db 61 KCLHLRLTGORAFCLIEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||

QY 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||
Db 121 LELGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||

RESULT 5
US-10-450-763-59832
; Sequence 59832, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 59832
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(246)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-59832

Query Match      21.4%; Score 164.5; DB 5; Length 246;
Best Local Similarity 60.3%; Pred. No. 1.3e-09;
Matches 47; Conservative 2; Mismatches 22; Indels 7; Gaps 4;

QY 64 HLRLTGORAFCLIEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 121
   |||||
Db 58 HSLSQQRACFLIGFLPWHSHVGLNECKVLLSGRSSQOMGEARRGMWEVFP 117
   |||||
QY 122 ELGRPEAGALQRLPQNS 139
   |||||
```

GenCore version 5.1.9
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2007, 15:01:16 ; Search time 38 Seconds
(without alignments)
379.803 Million cell updates/sec

Title: US-10-777-578-2

Perfect score: 770

Sequence: 1 MVVFTLLVLLLLLTALC.....LQRLPQPNALLACKACAGAY 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_80:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.5	10.8	254	H37081	proline/glycine be
2	75.5	9.8	483	A38560	nitrate transport
3	73	9.5	492	F83462	hypothetical prote
4	73	9.5	1163	F84669	probable chromosom
5	72.5	9.4	326	B46108	outer capsid prote
6	72.5	9.4	343	D75260	probable carboxydr
7	71.5	9.3	253	G30275	hypothetical prote
8	71	9.2	406	S76451	hypothetical prote
9	70.5	9.2	326	1 VGXRB	glycoprotein VP7 p
10	70.5	9.2	326	A44895	major glycoprotein
11	70	9.1	442	T06476	ppf-1 protein - ga
12	69.5	9.0	474	S75568	isochorismate synt
13	69.5	9.0	649	H64476	lipocate protein li
14	69	9.0	216	S70786	probable transcrip
15	69	9.0	253	A30509	cobalamin synthase
16	69	9.0	1212	T00332	hypothetical prote
17	68.5	8.9	532	T44788	phospho-prenol glu
18	68.5	8.9	597	E40201	artifact-warning s
19	68.5	8.9	1627	A21109	two-component hybr
20	68	8.8	309	AC1349	oligopeptide ABC t
21	68	8.8	618	T48193	hypothetical prote
22	68	8.8	1064	B86465	probable protein k
23	68	8.8	1932	T25525	hypothetical prote
24	67.5	8.8	260	C69893	probable enoyl-CoA
25	67.5	8.8	326	1 VGXR49	glycoprotein VP7 p
26	67.5	8.8	326	1 VGXR7H	glycoprotein VP7 p
27	67.5	8.8	326	1 VGXRMD	glycoprotein VP7 p
28	67.5	8.8	326	1 VGXRWA	glycoprotein VP7 p
29	67.5	8.8	326	2 S25546	outer capsid prote

30	67.5	8.8	1121	2	C82120	transcription regu
31	67	8.7	304	2	T52079	probable zinc fing
32	67	8.7	386	2	AD2649	ABC transporter, m
33	67	8.7	386	2	C97431	alpha-glucosides t
34	67	8.7	409	2	T03718	suppressor 2 prote
35	67	8.7	499	1	S66677	thioredoxin-disulf
36	67	8.7	551	2	D71969	1-lactate permease
37	67	8.7	1354	2	T48198	hypothetical prote
38	66.5	8.6	254	1	G64045	3-deoxy-manno-ctu
39	66.5	8.6	302	2	T31535	hypothetical prote
40	66.5	8.6	326	1	JQ1442	glycoprotein VP7 p
41	66.5	8.6	326	1	JQ1443	glycoprotein VP7 p
42	66.5	8.6	326	1	VGXRAB	glycoprotein VP7 p
43	66.5	8.6	326	1	VGXR2S	glycoprotein VP7 p
44	66.5	8.6	326	1	VGXR3S	glycoprotein VP7 p
45	66.5	8.6	326	1	VGXRDS	glycoprotein VP7 p

ALIGNMENTS

RESULT 1

H97081

proline/glycine betaine ABC transport system, ATPase component CAC1475 [imported] - C1
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: H97081

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Le
J. Bacteriol. 183, 4823-4838, 2001
J. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97081

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <KUR>

A:Cross-references: UNIPROT:Q97J11; UNIPARC:UPI00000CA1FA; GB:AE001437; PIDN:AAK79443.

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1475

Query Match 10.8%; Score 83.5; DB 2; Length 254;

Best Local Similarity 32.5%; Pred. No. 0.69;

Matches 37; Conservative 10; Mismatches 42; Indels 25; Gaps 6;

Qy 17 TALCKALSQLPYLYRPSQSLSFLLITDIIKKIDIQFLPLTG-----GKCLHLRLTQQ 70

Db 45 TLLKMINR-----LYEPDKGSIVLFNEDIKKIDVVKLRRSIGYVIOQVGLPHTMIANN 99

Qy 71 RAFCLLEFLPWCNGIIES-----LVGLE-NERK-----VLGGSSSQMGGEAR 111

Db 100 IA-TVPKLLKWDKREIKRIDEHLHLVGLFPNEFKKYPSPQLSGGQQQRIGLAR 152

RESULT 2

A38560

nitrate transport protein crna - Emericella nidulans

C:Species: Emericella nidulans, Aspergillus nidulans

C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 09-Jul-2004

C:Accession: A38560

R:Unkles, S.E.; Hawker, K.L.; Grieve, C.; Campbell, E.I.; Montague, P.; Kinghorn, J.R.

Proc. Natl. Acad. Sci. U.S.A. 88, 204-208, 1991

A:Title: crna encodes a nitrate transporter in Aspergillus nidulans.

A:Reference number: A38560; MUID:91095428; PMID:1986367

A:Accession: A38560

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-483 <UNK>

A:Cross-references: UNIPROT:P22152; UNIPARC:UPI00000178CA8; GB:M57647

C:Superfamily: nitrate transporter component

C:Keywords: transmembrane protein

Query Match 9.8%; Score 75.5; DB 2; Length 483;

```

Best Local Similarity 23.1%; Pred. No. 8.6;
Matches 31; Conservative 18; Mismatches 38; Indels 47; Gaps 7;

Qy      3 VFTLLVLLLLLALCKALSQSLTYLYRPOSSLSFLITDIKK-----ID 50
       :   |||:   |   |   |   |   |   |   |   |   |
Db      77 IIALATLLVRLICGCLCDRFGRPLVF-----ICLLLVGSIPTAMAGLTSPQGLIA 128
       :   |||:   |   |   |   |   |   |   |   |

Qy     51 IQVFELPLTGCKLHLRLTGORACILEFLP---WCNGIIE-SLVGLENERKVLSGSSQQ 106
       :   |||:   |   |   |   |   |   |   |   |
Db    129 LRFIFIGILG-----TFVFCQVWCTGFDFDKSIVGTANS---LAAG---- 165

Qy    107 MGEARRGMWEVEVPF 120
       :   |   |   |   |
Db   166 LGNAGGGITYFYVMP 179

```

RESULT 3
F83462
Hypothetical protein PA1474 [imported] - *Pseudomonas aeruginosa* (strain PAO1)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text_change 09-Jul-2004
R:Accession: F83462
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
L.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: F83462
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-492 <STO>
A:Cross-references: UNIPROT:Q913N8; UNIPARC:UP100000C5395; GB:AE004576; GB:AE004091; NID
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA1474

Query Match	9.5%; Score 73; DB 2; Length 492;
Best Local Similarity	32.2%; Pred.No.16;
Matches	39; Conservative 15; Mismatches 39; Indels 28; Gaps 7;
QY	10 LLLLLLLTALC-KALSQSLPVTLYRPOSSLSFLLITDIKKIDIQFLPLTG---GKCLHL 65 : : : : : : : :
Db	124 VLNLNLTPLAGSKLSDSQAL--PVGSLLTALVKNQQID---FLPSGRLDQLALGQ 178 : : : : : : : :
QY	66 RLTCQRA-----FCILEBFLPCWNGIIESL-----VGLENERKVLSGGSSQQMGE 109 : : : : : : : :
Db	179 QLAGSARQASLEGLFAALQCLGNCRNDLPDSLRQAVERLFAGLPEAQKM---GD SKLLGQ 235 : : : : : : : :
QY	110 A 110
Db	236 A 236

RESULT 4
F84669
probable chromosome associated protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 31-Dec-2004
C:Accession: F84669
C:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84669
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1163 <STO>
A:Cross-references: UNIPROT:Q9SHT1; UNIPARC:UPI0000178BFD; GB:AE002093; NID:96598752; PT:
C:Genetics:
A:Gene: At2q27170

A:Map position: 2

Query Match 9.5%; Score 73; DB 2; Length 1163;
Best Local Similarity 25.5%; Pred. No. 38;
Matches 28; Conservative 27; Mismatches 37; Indels 18; Gaps 5;

QY 24 SQSPYTLRPPQSSLSFLIIDIKIDIOYFLPTGGKCLH-----LRL---TGORAF 73
DB 131 SRANPYVVOGKTASLTMKDIEKLDL---LKEIGTRVYERRRESIRIMQETGNKRK 187
QY 74 CILBFLPWCNGIIBSLVGLNERKVLGGSSQQMGGEARRGMEWEVPLEL 123
DB 188 QIIIEVHYLD---EFLRELDEEKELR--KYQQLDKQKSLEYTIYDKEL 232

RESULT 5
B46108
outer capsid protein VP7 - human rotavirus (strain I321)
C:Species: human rotavirus
C:Date: 07-Apr-1994 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: B46108
R:Das, M.; Dunn, S.J.; Woode, G.N.; Greenberg, H.B.; Rao, C.D.
Virolgy 194, 374-379, 1993
A:Title: Both surface proteins (VP4 and VP7) of an asymptomatic neonatal rotavirus str
A:Reference number: A46108; MUID:93242771; PMID:8386681
A:Contents: I321, serotype 10
A:Accession: B46108
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-326 <DAS>
A:Cross-references: UNIPROT:Q86188; UNIPARC:UPI00001786EF; GB:L07658; NID:g310326; PII
A:Note: sequence extracted from NCBI backbone (NCBIN:130701, NCBIP:130693)
A:Note: the sequence in GenBank entry R01VP7B, release 109, (PID:g310325) has the cod
C:Superfamily: rotavirus glycoprotein VP7

Query Match 9.4%; Score 72.5; DB 2; Length 326;
Best Local Similarity 30.0%; Pred. No. 12;
Matches 18; Conservative 14; Mismatches 17; Indels 11; Gaps 2;

QY 5 TLLAVLLELLLTALCKALQSLSPLTYLRPQSSLSFLITDIIKKIDQ-----YFLPLTG 59
DB 8 TFLIVLILILFNLYLKSIITRMDDYIYK-----FLLIWITIASIIVNPQNYGINLPITG 61

RESULT 6
D75260
probable carboxylate kinase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: D75260
R:White, O.; Eلسen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: D75260
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-343 <WHI>
A:Cross-references: UNIPROT:Q9RRD9; UNIPARC:UPI00000D3FDE; GB:AE002084; GB:AE000513;
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2553
A:Map position: 1
C:Superfamily: ribokinase

Query Match 9.4%; Score 72.5; DB 2; Length 343;
Best Local Similarity 30.5%; Pred. No. 12;
Matches 29; Conservative 12; Mismatches 33; Indels 21; Gaps 3;

QY 65 LRLTGORAFCLIEFLPWCNGIIESLVGLENERKVLGGSSQQMGGEARRGMEWEVFPLELG 124

Db 82 LRAEGVRAE-VLASAAHPTGVILALIDRRGQAMLTG-----QGADWELLPEELP 130

QY 125 RP-----EAGLQRLPQNSALLACRCAGA 149

Db 131 RDVLSAGHLHTAWSLFRDPRAAALAEARIAKA 165

RESULT 7

G90275

Hypothetical protein SSO1215 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: G90275

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sassen, C.W.; Van der Oost, J.
submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90275

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-253 <KUR>

A:Cross-references: UNIPROT:Q97YU2; UNIPARC:UPI00000643AE; GB:AE006641; NID:gi13814412; F

C:Genetics:

A:Gene: SSO1215

Query Match 9.3%; Score 71.5; DB 2; Length 253;

Best Local Similarity 32.1%; Pred. No. 11;

Matches 26; Conservative 14; Mismatches 34; Indels 7; Gaps 3;

QY 2 VVFTLLVLLLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQFLPLTGK 61

Db 96 VVFTKLGVLMLVLSKSYAK-----IPEENFSTLKIKYDLMEEVKRIRI-VSLPLLYSK 149

QY 62 CLHLRLTQO-RAPCILEFLPW 81

Db 150 ILWLRCKVEKIRELSILKTKDW 170

RESULT 8

S76451

Hypothetical protein - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.

A:Reference number: S74322; MUID:97061201; PMID:9905231

A:Accession: S76451

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-406 <KAN>

A:Cross-references: UNIPROT:P74479; UNIPARC:UPI000000C1038; EMBL:D90915; GB:AB001339; NID

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

Query Match 9.2%; Score 71; DB 2; Length 406;

Best Local Similarity 23.3%; Pred. No. 21;

Matches 31; Conservative 22; Mismatches 34; Indels 46; Gaps 6;

QY 11 LLLLLLTALCKALSQSLPYLYRQSSLSFLII-----TDIKKIDIQF--LPLTGG 60

Db 41 LLLVLAVCLFGATNSMAGWLY-VISGISFALLITAAIILPWSLQQLERFPQPVSAG 99

QY 61 KCHLRLTQORAFCLIFLPCWNGHIESLVGLNERKVLSSQSQMGAEARGME-WEYF 119

Db 100 EDLTITLR-----LHNHKK-----EAKNLLQWVDVL 125

QY 120 PLELGRPEAGALQ 132

Db 126 PTGLRPOGEAVE 138

RESULT 9

VGXRBB

glycoprotein VP7 precursor - bovine rotavirus A (serotype 10 strain B223)

C:Species: bovine rotavirus A

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999

C:Accession: A38510

R:Xu, L.; Harbour, D.; McCrae, M.A.

J. Gen. Virol. 72, 177-180, 1991

A:Title: Sequence of the gene encoding the major neutralization antigen (VP7) of serot

A:Reference number: A38510; MUID:91116309; PMID:1703560

A:Accession: A38510

A:Molecule type: genomic RNA

A:Residues: 1-326 <XUL>

A:Cross-references: UNIPARC:UPI00000619B6; EMBL:X52650; NID:G61512; PIDN:CAA36875.1; P

C:Genetics:

A:Map position: segment 9

C:Superfamily: rotavirus glycoprotein VP7

C:Keywords: coat protein; Glycoprotein; transmembrane protein

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-326/Product: glycoprotein VP7 #status predicted <GPV>

F:32-48/Region: hydrophobic #status predicted

F:69,145/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.2%; Score 70.5; DB 1; Length 326;

Best Local Similarity 30.0%; Pred. No. 19;

Matches 18; Conservative 14; Mismatches 17; Indels 11; Gaps 2;

QY 5 TLLVLLVLLLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQ-----YFLPLTG 59

Db 8 TFLYLISIIILLNVLKSLITRMMDYIYK-----FLLIVTITSIVNAQNYGINLPTG 61

RESULT 10

A44895

major glycoprotein VP7 - bovine rotavirus

C:Species: bovine rotavirus

C>Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A44895

R:Huang, J.A.; Nagesha, H.S.; Snodgrass, D.R.; Holmes, I.H.

J. Clin. Microbiol. 30, 85-92, 1992

A:Title: Molecular and serological analyses of two bovine rotaviruses (B-11 and B-60)

A:Reference number: A44895; MUID:92129681; PMID:1310336

A:Accession: A44895

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-326 <HUA>

A:Cross-references: UNIPROT:Q86192; UNIPARC:UPI000000F45B0; GB:M64679; NID:G333829; PID

A:Experimental source: strain B-11

A>Note: sequence extracted from NCBI backbone (NCBIN:78093, NCBIP:78154)

C:Superfamily: rotavirus glycoprotein VP7

C:Keywords: glycoprotein

Query Match 9.2%; Score 70.5; DB 2; Length 326;

Best Local Similarity 30.0%; Pred. No. 19;

Matches 18; Conservative 14; Mismatches 17; Indels 11; Gaps 2;

QY 5 TLLVLLVLLLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQ-----YFLPLTG 59

Db 8 TFLYLISIIILLNVLKSLITRMMDYIYK-----FLLIVTITSIVNAQNYGINLPTG 61

RESULT 11

T06476

ppf-1 protein - garden pea

C:Species: Pisum sativum (garden pea)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

C:Accession: T06476

R;Zhu, Y.; Zhang, Y.; Luo, J.; Davies, P.J.; Ho, D.T.H.
Gene 208, 1-6, 1998
A;Title: PPF-1, a post-floral-specific gene expressed in short-day-grown G2 pea, may be
A;Reference number: Z15705; MUID:98147997; PMID:9479033
A;Accession: T06476
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-442 <ZHU>
A;Cross-references: UNIPROT:Q9FY06; UNIPARC:UPI00000ABB7B; EMBL:Y12618; NID:94218522; P
A;Experimental source: cv. G2
C;Genetics:
A;Gene: ppf-1
F;93-330/Domain: stage III sporulation protein homology <SPOR>

Query Match 9.1%; Score 70; DB 2; Length 442;
Best Local Similarity 25.2%; Pred. No. 29;
Matches 35; Conservative 24; Mismatches 50; Indels 30; Gaps 7;

QY 13 LLLLTALCKALSQSLPVTLYRPOQSLGFL-LITDIKKIDIOYFLPLTGGKCLHLRLTGQR 71

Db 122 IILITVIVKA--ATLPLTKQVESTLAMONLQPKIKAIQERY-----AGNQERIQLETSR 174

QY 72 AF-----CI--LEFLPWCNGIIESLVGLNERKV-----LSGGSSQQMGEA 110

Db 175 LYTQAGVNPAGCLPTLATIPVMIGLYCALSNVANEGLLTGFLWIPSLGGPTTSIAARQS 234

QY 111 RRGMEWEVFPLELGRPRAG 129

Db 235 GSGISW-LFPFVDGHPPLG 252

RESULT 12

S75568

isochorismate synthase (EC 5.4.99.6) - *Synechocystis* sp. (strain PCC 6803)

N;Alternate names: protein slr0817

C;Species: *Synechocystis* sp.

A;Variety: PCC 6803

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C;Accession: S75568

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S75568

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-474 <KAN>

A;Cross-references: UNIPARC:UPI00000D3511; EMBL:D90911; GB:AB001339; NID:gl653083; PIDN:

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Gene: entC

C;Superfamily: isochorismate synthase

C;Keywords: intramolecular transferase; isomerase

Query Match 9.0%; Score 69.5; DB 1; Length 474;
Best Local Similarity 28.6%; Pred. No. 34;
Matches 42; Conservative 16; Mismatches 58; Indels 31; Gaps 8;

QY 19 LCKALSQSLPVTLYRPOQSLGFLITDIKKIDIOYFLPLTGGKCLHLRLTGQAFCLILEF 78

Db 215 LSKAIASLEIEI---AQRSLKVLA--TALDLYGSLNVAHCLQ-RLRQYGDVYL-- 266

QY 79 LPWCNGIIESLVGLNER-----KVLSSGSSQQM---GEARRGMEWEVFPLELG 124

Db 267 FSWGNGQDCFGASPERLLSLHNLQVTDALAGSAPRDVDQGRDLQGLLHNPKELR 326

QY 125 RPEA---GALQRL-----PQNSALL 142

Db 327 EHQAVLDYLLQRLRALGSLSPQASSLKL 353

RESULT 13

H64476

lipote protein ligase (EC 3.4.21.-) *Lon* - *Methanococcus jannaschii*

N;Alternate names: Alp-dependent proteinase lon homolog

C;Species: *Methanococcus jannaschii*

C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004

C;Accession: H64476

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, J
ison, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,

A;Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: H64476

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-649 <BUL>

A;Cross-references: UNIPROT:Q58812; UNIPARC:UPI000012E7EC; GB:U67582; GB:L77117; NID:

C;Genetics:

A;Map Position: FOR1378485-1380434

A;Superfamily: Lon-type protease

C;Keywords: ATP; hydrolase; nucleotide binding; P-loop; serine proteinase

F;6-89/Domain: *Methanococcus* endopeptidase La homolog P-loop-containing homology <MLA

F;47-54/Region: nucleotide-binding motif A (P-loop)

F;238-243/Region: nucleotide-binding motif B

F;550/Active site: Ser #status predicted

Query Match 9.0%; Score 69.5; DB 1; Length 649;

Best Local Similarity 27.6%; Pred. No. 47;

Matches 42; Conservative 24; Mismatches 53; Indels 33; Gaps 10;

QY 2 VVFTLLLVLL--LLLLLTALCKALSQSLPVTLYRPOQSLGFLITDIKKIDIOYFLPLTG 59

Db 143 VTITLVIVFIFGVIIITSLMGASRAMNNL-NPM-DLXPVLYECK-----RLPVR 194

QY 60 GKCLHL-RLTGQRAFCILFPLPWCNGIIESLVGLNERKVLSSGSSQQMGEGRMG--- 114

Db 195 ASAYNTRLLGDIKHCPGLGRP-----PLCTPPHKRIILGA----IHEAHRGILYVD 242

QY 115 EWEVFPLELGRPRAGALQ--RLP-----QPNSA 140

Db 243 EIKTMPLEVDQYIITALQDKQLPISGRNPNS 274

RESULT 14

S70786

probable transcription regulator *csgD* - *Escherichia coli* (strain K-12)

C;Species: *Escherichia coli*

C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 09-Jul-2004

C;Accession: S70786; E64846

R;Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A;Title: Expression of two *csg* operons is required for production of fibronectin- and

A;Reference number: S70783; MUID:96414468; PMID:8817489

A;Accession: S70786

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-216 <HAM>

A;Cross-references: UNIPROT:P52106; UNIPARC:UPI000012852F; EMBL:X90754; NID:gl147558;

A;Experimental source: strain K12, substrain MC4100

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: E64846

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-216 <BLAT>

A;Cross-references: UNIPARC:UPI000012852F; GB:AE000205; GB:U000096; NID:gl787265; PIDN

A;Experimental source: strain K-12, substrain MG1655

```

Query Match          9.0%; Score 69; DB 2; Length 253;
Best Local Similarity 23.0%; Pred.No. 20;
Matches 35; Conservative 32; Mismatches 55; Indels 30; Gaps 7;

Qy      7   LLVLLLLLLLTAACKALQSPLVTLYRPQSSFLLTIDIKYDIQYFLPT-----58
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      113 LLVVYSIQVALK-----LDPSLY----TIFLISSNVLSWSLSYLITSTSPESN 163
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy      59   GGKCLHLRLTGORAFCIIELPWCN--GIIESVLGLENERKVKLS--GGSSQGAEARRM 114
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      164   LGRIFHDKLAGKSTILLLELIPISLVYNVVIIFYIMYMKICGSLGGSSGDAGASITL 223
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy      115   EWEVFPEELGRPEGALQRLPQNALLACRC 146
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      224   S---FPIFLITDEITNL-----NYSLLSILC 246
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

```

Search completed: January 10, 2007, 15:06:43
Job time : 40 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2007, 15:06:17 ; Search time 50 Seconds
(without alignments)
262.592 Million cell updates/sec

Title: US-10-777-578-2

Perfect score: 770

Sequence: 1 MVVFTLLVLLVLLLTALC.....LQLPQPNALLACRCAGAY 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /EMC Celerra_SIDS3/ptodata/2/iaa/5 COMB.pdp:*
 - 2: /EMC Celerra_SIDS3/ptodata/2/iaa/6 COMB.pdp:*
 - 3: /EMC Celerra_SIDS3/ptodata/2/iaa/7 COMB.pdp:*
 - 4: /EMC Celerra_SIDS3/ptodata/2/iaa/H COMB.pdp:*
 - 5: /EMC Celerra_SIDS3/ptodata/2/iaa/PCITUS COMB.pdp:*
 - 6: /EMC Celerra_SIDS3/ptodata/2/iaa/RE COMB.pdp:*
 - 7: /EMC Celerra_SIDS3/ptodata/2/iaa/backfiles1.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	770	100.0	150	2	US-09-893-737-230
2	75.5	9.8	415	2	US-09-540-236-2600
3	73.5	9.5	256	2	US-09-949-016-8503
4	73	9.5	247	2	US-09-675-305-6
5	73	9.5	247	2	US-10-200-344-6
6	73	9.5	350	2	US-09-675-305-12
7	73	9.5	350	2	US-10-200-344-12
8	73	9.5	437	2	US-09-675-305-10
9	73	9.5	437	2	US-10-200-344-10
10	73	9.5	531	2	US-09-252-991A-27660
11	70.5	9.2	326	1	US-08-671-978A-7
12	70.5	9.2	326	2	US-09-232-338-7
13	69	9.0	109	2	US-09-688-017-310
14	69	9.0	309	3	US-10-770-127-60
15	68	8.8	183	2	US-09-893-737-250
16	68	8.8	263	2	US-09-710-279-332
17	68	8.8	274	2	US-09-134-001C-3921
18	68	8.8	419	2	US-09-543-681A-4221
19	68	8.8	462	2	US-09-328-352-7472
20	67.5	8.8	326	1	US-07-603-133B-27
21	67.5	8.8	670	1	US-08-366-547-2
22	67	8.7	497	2	US-09-949-016-6343
23	67	8.7	502	1	US-08-960-022-10
24	67	8.7	549	1	US-08-500-635A-12
25	67	8.7	549	2	US-09-167-151-12
26	67	8.7	549	2	US-09-842-347A-12

27	66.5	8.6	326	2	US-08-089-397A-11	Sequence 11, Appl
28	66.5	8.6	326	7	5395759-2	Patent No. 5395759
29	66.5	8.6	503	2	US-09-252-991A-30682	Sequence 30682, A
30	66.5	8.6	748	2	US-09-252-991A-21896	Sequence 21896, A
31	66.5	8.6	4861	2	US-09-919-497-70	Sequence 70, Appl
32	66	8.6	306	3	US-10-770-127-62	Sequence 62, Appl
33	65.5	8.5	254	1	US-08-236-918A-4	Sequence 4, Appl
34	65.5	8.5	254	2	US-09-150-864A-4	Sequence 4, Appl
35	65.5	8.5	267	2	US-09-248-796A-21552	Sequence 14, Appl
36	65.5	8.5	326	2	US-08-089-397A-14	Sequence 21552, A
37	65.5	8.5	530	2	US-09-252-991A-23552	Sequence 23552, A
38	65	8.4	175	2	US-09-252-991A-23562	Sequence 23562, A
39	65	8.4	285	2	US-09-252-991A-25160	Sequence 25160, A
40	65	8.4	319	3	US-10-770-127-51	Sequence 51, Appl
41	65	8.4	386	1	US-08-134-012-3	Sequence 3, Appl
42	65	8.4	386	1	US-08-520-519-3	Sequence 3, Appl
43	65	8.4	386	2	US-09-039-798-3	Sequence 3, Appl
44	65	8.4	423	2	US-09-489-039A-9464	Sequence 9464, Ap
45	64.5	8.4	293	2	US-09-902-540-12489	Sequence 12489, A

ALIGNMENTS

RESULT 1

US-09-893-737-230
; Sequence 230, Application US/09893737
; Patent No. 5395759-2
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 230
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-230

Query Match 100.0%; Score 770; DB 2; Length 150;

Best Local Similarity 100.0%; Pred. No. 1.5e-85;

Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVVFTLLVLLVLLLTALCKALSQSLPYTYLRQSSLSFLITDKKIDIQYFLPTGG 60

Db 1 MVVFTLLVLLVLLLTALCKALSQSLPYTYLRQSSLSFLITDKKIDIQYFLPTGG 60

Qy 61 KCLHRLTGTGAFACILEFLPWCNGIIESLVGLENERKVLSCGSSQOMGEARGMEWEVFP 120

Db 61 KCLHRLTGTGAFACILEFLPWCNGIIESLVGLENERKVLSCGSSQOMGEARGMEWEVFP 120

Qy 121 LELGRPAGALQRLPQPNALLACRCAGAY 150

Db 121 LELGRPAGALQRLPQPNALLACRCAGAY 150

RESULT 2

US-09-540-236-2600

; Sequence 2600, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CAT.

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04


```

; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2600
; LENGTH: 415
; TYPE: PRT
; ORGANISM: M.cattarrhalis
US-09-540-236-2600

Query Match
Best Local Similarity 24.6%; Pred. No. 1.9; DB 2; Length 415;
Matches 32; Conservative 23; Mismatches 44; Indels 31; Gaps 5;

QY 5 TLLTLLTLLTALCKALQSPLYTYRPOSSLSFL-----LITDIKKIDIQYFLPLT 58
Db 43 TLLTLLTLLTALCKALQSPLYTYRPOSSLSFL-----LITDIKKIDIQYFLPLT 58
QY 59 GKGKHLRLTGORAFCLILEFLPWNG-----IISLVGLNERKVLGGSSQQMGGEARR 114
Db 101 -----LRVYQRTTHNVVFLMMVTALGLLILTAMFG-----DVGING-----SRRWL 141
QY 115 EWEVFPLELG 124
Db 142 DLGIFNLQAG 151

RESULT 3
US-09-949-016-8503
; Sequence 8503, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8503
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8503

Query Match
Best Local Similarity 22.9%; Pred. No. 1.7; DB 2; Length 256;
Matches 32; Conservative 12; Mismatches 45; Indels 51; Gaps 4;

QY 12 LLLLTALCKALQSPLYTYRPOSSLSFLITDIKKIDIQYFLPLTGGKCLHLRLTGQR 71
Db 13 LLLVILTEPDYLNHHPCTCKPQENMG-----RVFLTGEK 48
QY 72 AFCILEFLPWNGIIESL-----VGLNERKV-----LSGGSSQQMG 108
Db 49 ANSILKRYPRANGFPFEEIROGNIERECKEEFCTFEAREAFENNEKTFEWSYTKAQGG 108
QY 109 EARGMEW-----EVPFLELG 124
Db 109 ESNRGSDWFQFYLTFFLIFG 128

RESULT 4
US-09-675-305-6
; Sequence 6, Application US/09675305
; Patent No. 6411531
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
US-09-675-305-6

; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NO. 6411531 Human Carboxypeptidases and
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/09/675,305
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: US 60/156,685
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 247
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-675-305-6

Query Match
Best Local Similarity 29.5%; Pred. No. 1.9; DB 2; Length 247;
Matches 31; Conservative 16; Mismatches 30; Indels 28; Gaps 6;

QY 21 KALSQSLPYTYRPOSSLSFL-----LITDIKKIDIQYFLPLTGGKCLHLRLTGORAFCLILE 77
Db 57 KKISYQLKVDLWQP--SSISYVSEGTVDV-----HIPQNGSRA--LLA 96
QY 78 FLPCWNGIIESLVGLNERKVLGGSSQQMGGEARR---GMEWEVVF 119
Db 97 FLOEAN--IQYKVLIEDLQTKLEKSSLTQNRNRRSLSGYNYEVY 139

RESULT 5
US-10-200-344-6
; Sequence 6, Application US/10200344
; Patent No. 6780640
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6780640el Human Carboxypeptidases and
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 247
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-200-344-6

Query Match
Best Local Similarity 29.5%; Pred. No. 1.9; DB 2; Length 247;
Matches 31; Conservative 16; Mismatches 30; Indels 28; Gaps 6;

QY 21 KALSQSLPYTYRPOSSLSFL-----LITDIKKIDIQYFLPLTGGKCLHLRLTGORAFCLILE 77
Db 57 KKISYQLKVDLWQP--SSISYVSEGTVDV-----HIPQNGSRA--LLA 96
QY 78 FLPCWNGIIESLVGLNERKVLGGSSQQMGGEARR---GMEWEVVF 119
Db 97 FLOEAN--IQYKVLIEDLQTKLEKSSLTQNRNRRSLSGYNYEVY 139

RESULT 6
US-09-675-305-12
```

```

; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2600
; LENGTH: 415
; TYPE: PRT
; ORGANISM: M.cattarrhalis
US-09-540-236-2600

Query Match
Best Local Similarity 24.6%; Pred. No. 1.9; DB 2; Length 415;
Matches 32; Conservative 23; Mismatches 44; Indels 31; Gaps 5;

QY 5 TLLTLLTLLTALCKALQSPLYTYRPOSSLSFL-----LITDIKKIDIQYFLPLT 58
Db 43 TLLTLLTLLTALCKALQSPLYTYRPOSSLSFL-----LITDIKKIDIQYFLPLT 58
QY 59 GKGKHLRLTGORAFCLILEFLPWNG-----IISLVGLNERKVLGGSSQQMGGEARR 114
Db 101 -----LRVYQRTTHNVVFLMMVTALGLLILTAMFG-----DVGING-----SRRWL 141
QY 115 EWEVFPLELG 124
Db 142 DLGIFNLQAG 151

RESULT 3
US-09-949-016-8503
; Sequence 8503, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8503
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8503

Query Match
Best Local Similarity 22.9%; Pred. No. 1.7; DB 2; Length 256;
Matches 32; Conservative 12; Mismatches 45; Indels 51; Gaps 4;

QY 12 LLLLTALCKALQSPLYTYRPOSSLSFLITDIKKIDIQYFLPLTGGKCLHLRLTGQR 71
Db 13 LLLVILTEPDYLNHHPCTCKPQENMG-----RVFLTGEK 48
QY 72 AFCILEFLPWNGIIESL-----VGLNERKV-----LSGGSSQQMG 108
Db 49 ANSILKRYPRANGFPFEEIROGNIERECKEEFCTFEAREAFENNEKTFEWSYTKAQGG 108
QY 109 EARGMEW-----EVPFLELG 124
Db 109 ESNRGSDWFQFYLTFFLIFG 128

RESULT 4
US-09-675-305-6
; Sequence 6, Application US/09675305
; Patent No. 6411531
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
US-09-675-305-6

; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NO. 6411531 Human Carboxypeptidases and
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/09/675,305
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: US 60/156,685
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 247
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-675-305-6

Query Match
Best Local Similarity 29.5%; Pred. No. 1.9; DB 2; Length 247;
Matches 31; Conservative 16; Mismatches 30; Indels 28; Gaps 6;

QY 21 KALSQSLPYTYRPOSSLSFL-----LITDIKKIDIQYFLPLTGGKCLHLRLTGORAFCLILE 77
Db 57 KKISYQLKVDLWQP--SSISYVSEGTVDV-----HIPQNGSRA--LLA 96
QY 78 FLPCWNGIIESLVGLNERKVLGGSSQQMGGEARR---GMEWEVVF 119
Db 97 FLOEAN--IQYKVLIEDLQTKLEKSSLTQNRNRRSLSGYNYEVY 139

RESULT 5
US-10-200-344-6
; Sequence 6, Application US/10200344
; Patent No. 6780640
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6780640el Human Carboxypeptidases and
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 247
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-200-344-6

Query Match
Best Local Similarity 29.5%; Pred. No. 1.9; DB 2; Length 247;
Matches 31; Conservative 16; Mismatches 30; Indels 28; Gaps 6;

QY 21 KALSQSLPYTYRPOSSLSFL-----LITDIKKIDIQYFLPLTGGKCLHLRLTGORAFCLILE 77
Db 57 KKISYQLKVDLWQP--SSISYVSEGTVDV-----HIPQNGSRA--LLA 96
QY 78 FLPCWNGIIESLVGLNERKVLGGSSQQMGGEARR---GMEWEVVF 119
Db 97 FLOEAN--IQYKVLIEDLQTKLEKSSLTQNRNRRSLSGYNYEVY 139

RESULT 6
US-09-675-305-12
```

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	770	100.0		150	5	AAU83189	Aau83189 Novel sec
2	770	100.0		150	5	AAW50708	Aam50708 Human sec
3	164.5	21.4		246	4	ABG29473	Abg29473 Novel hum
4	134	17.4		635	4	ABG19492	Abg19492 Novel hum
5	134	17.4		2134	4	ABG27624	Abg27624 Novel hum
6	132.5	17.2		138	4	ABG05801	Abg05801 Novel hum
7	124.5	16.2		313	4	ABB12469	Abb12469 Human bon
8	96	12.5		216	8	ADU02421	Adu02421 Novel hum
9	95.5	12.4		87	4	ABG23765	Abg23765 Novel hum
10	93	12.1		87	4	ABG12819	Abg12819 Novel hum
11	84	10.9		263	4	ABG24127	Abg24127 Novel hum
12	84	10.9		263	4	ABG07958	Abg07958 Novel hum
13	79.5	10.3		459	8	ADP30331	Adp30331 Human sec
14	77.5	10.1		322	7	ADP36760	Adp36760 Rhesus ro
15	75.5	9.8		391	6	ABU35437	Abu35437 Protein e
16	75.5	9.8		415	8	ADU04914	Adi04914 M. catarr
17	75	9.7		527	8	ADN26981	Adn26981 Bacterial
18	74	9.6		137	4	AAU48296	Aau48296 Propionib
19	74	9.6		137	6	ABM44815	Abm44815 Propionib
20	74	9.6		307	8	ADH42051	Adh42051 Novel hum
21	74	9.6		1105	7	ABM90228	Abm90228 Rice abio
22	74	9.6		1218	7	ABM88074	Abm88074 Rice abio
23	73.5	9.5		414	7	ADP03569	Adp03569 Human GPC

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 59832; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 246 AA;
XX Query Match 21.4%; Score 164.5; DB 4; Length 246;
XX Best Local Similarity 60.3%; Pred. No. 9.7e-10;
XX Matches 47; Conservative 2; Mismatches 22; Indels 7; Gaps 4;
OY 64 HRLTGQRAFCLILEPLPWCNGIIESLVGLENERKV-LSGSSQQMGE-ARRGMWEVFPPL 121
Db 58 HSLSGQRAFCLILGFLPWHRRIRSHVGLNECKVLLSGRSSQMGEPXGRFSPXVGPL 117
OY 122 ELGRPEAGALQRLPQPN 139
Db 118 G-GR----ALLQLPQPN 130
RESULT 4
ABG19492
ID ABG19492 standard; protein; 635 AA.
XX AC ABG19492;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #19483.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI; 2001-639362/73.
XX N-PSDB; AAS83679.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 49851; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 635 AA;
XX Query Match 17.4%; Score 134; DB 4; Length 635;
XX Best Local Similarity 48.1%; Pred. No. 8.9e-06;
XX Matches 38; Conservative 8; Mismatches 27; Indels 6; Gaps 3;
OY 76 LEFLPWCNGIIESLVGLENERKV-LSGSSQQMGEARRGMWEVFPPLGLRPEAGALQRL 134
Db 399 LGFLPWCNTRRIGSHVGLNECQVLLSGSSQQMKPEGRWSGKVSP-GVGPLSSWTLLQL 457
OY 135 PQPNLSAL----LACRCAGA 149
Db 458 PWPNSALFCRWMACLSAGS 476
RESULT 5
ABG27624
ID ABG27624 standard; protein; 2134 AA.
XX AC ABG27624;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27615.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX

XX 30-MAR-2001; 2001WO-US010472.
 XX
 XX 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 PR 30-NOV-2000; 2000US-0250583P.
 XX (HYSE-) HYSQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT, Ford JE, Boyle BJ;
 PI WPI; 2001-626375/72.
 XX
 XX New bone marrow-expressed nucleic acids and polypeptides, useful for
 PT diagnosis, treatment of inflammatory, autoimmune, neurological, cancer
 PT and increasing hematopoiesis, stem cell survival and bone growth and
 PT remodeling.
 XX
 XX Claim 10; Page 359-360; 380pp; English.
 PS
 CC The present invention relates to bone marrow expressed polynucleotides
 CC and proteins. These sequences can be used in the treatment of
 CC inflammatory conditions (eg arthritis, Crohn's disease), cancer, central
 CC and peripheral nervous system diseases and neuropathies, such as
 CC Alzheimer's, Parkinson's and Huntington's diseases, spinal cord
 CC disorders, head trauma, cerebrovascular diseases, myeloid and lymphoid
 CC cell disorders, platelet disorders, stem cell disorders, bone
 CC degenerative disorders, autoimmune disorders, for example multiple
 CC sclerosis, diabetes and arthritis, viral and bacterial infections, a
 CC allergies and blood coagulation disorders. The present sequence is a
 CC protein of the invention
 XX
 XX Sequence 313 AA;
 SQ
 Query Match 16.2%; Score 124.5; DB 4; Length 313;
 Best Local Similarity 45.9%; Pred. NO. 4.2e-05;
 Matches 39; Conservative 3; Mismatches 28; Indels 15; Gaps 5;
 QY 60 GKCLHLRLTGQAFCLFELPMCGHIESLVGLNERKV-LSGSSQOMGEARRGMEVEV 118
 Db 228 GTWLSRRTW--SFLYGGFLPMCSGRIGRVLGNECKVSLSGSSQPMGEPEG--RW-- 281
 QY 119 FPELGRPEAGALQRLPQNSALLA 143
 Db 282 -----SSPEVGP---LASFGSPLIA 298
 RESULT 8
 ID ADU02421
 XX ADU02421 standard; protein; 216 AA.
 AC ADU02421;
 XX
 XX 27-JAN-2005 (first entry)
 DT
 XX Novel human polypeptide seqid 888.
 DE
 XX cytostatic; antipsoriatic; antiinflammatory; gene therapy; Nanodisc;
 KW proliferative disorder; inflammatory disorder; immune disorder;
 KW metabolic disorder; bone disorder; CNS disorder; cancer; psoriasis;
 KW ulcerative colitis; human.
 XX
 OS Homo sapiens.
 XX
 XX WO2004093804-A2.
 PN
 XX 04-NOV-2004.
 PD
 XX 19-APR-2004; 2004WO-US012047.
 PF
 XX 18-APR-2003; 2003US-0463708P.
 PR 18-APR-2003; 2003US-0463732P.
 PR 02-MAY-2003; 2003US-0467199P.

PR 02-MAY-2003; 2003US-0467230P.
 PR 19-MAY-2003; 2003US-0471306P.
 PR 19-MAY-2003; 2003US-0471336P.
 PR 08-JUL-2003; 2003US-0485223P.
 PR 08-JUL-2003; 2003US-0485224P.
 PR 14-JUL-2003; 2003US-0486446P.
 PR 14-JUL-2003; 2003US-0486480P.
 PR 08-AUG-2003; 2003US-0493573P.
 PR 08-AUG-2003; 2003US-0493577P.
 PR 08-SEP-2003; 2003US-0505059P.
 XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
 PA
 XX Lee E, Hestir K, Chu K, Masuoka L, Williams LT;
 PI WPI; 2004-775861/76.
 XX N-PSDB; ADU01689.
 DR
 XX New first nucleic acid molecule comprising a polynucleotide sequence
 PT given in the specification, useful in preparing a composition for
 PT diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.
 PT
 PS Claim 14; SEQ ID NO 888; 291pp; English.
 XX
 CC The invention describes a new first nucleic acid molecule comprising a
 CC polynucleotide sequence given in the specification. Also described are:
 CC an animal injected with the nucleic acid molecule; a second nucleic acid
 CC molecule comprising a second polynucleotide sequence that is at least
 CC about 70, 80, 90 or 95% homologous to the first nucleic acid molecule or
 CC that hybridises to the first polynucleotide sequence under high
 CC stringency conditions; a vector comprising the nucleic acid molecule and
 CC a promoter that drives the expression of the nucleic acid molecule; a
 CC host cell transformed, transfected, transduced or infected with the
 CC nucleic acid molecule; a nucleic acid composition comprising a carrier or
 CC a buffer and one or more compositions comprising the nucleic acid
 CC molecule, vector or host cell; a substantially purified polypeptide; an
 CC animal injected with the polypeptide; a polypeptide composition
 CC comprising the polypeptide molecule and a carrier or buffer; a cell
 CC culture medium comprising the polypeptide or transfected cells
 CC transfected with the polynucleotide; making a transformed, transfected,
 CC transduced, or infected host cell; synthesising Nanodiscs simultaneously
 CC and for synthesising a series of simultaneously-synthesised Nanodiscs
 CC sequentially utilising a dynamic system; preparing a hydrophobic protein
 CC for determination of crystal structure; immunising a non-human animal;
 CC screening for modulators of hydrophobic protein activity; a diagnostic
 CC kit; determining the presence of the nucleic acid molecule or its
 CC complement; determining the presence of an antibody to the polypeptide in
 CC a sample; an antibody specifically recognising, binding to or modulating
 CC the biological activity of at least one polypeptide encoded by a nucleic
 CC acid molecule or its biologically active fragment; an antibody
 CC composition comprising the antibody and a carrier; a bacteriophage, where
 CC the antibody is displayed on the bacteriophage; a bacterial cell
 CC comprising the bacteriophage; a non-human animal injected with the
 CC antibody composition; a host cell that secretes the antibody; making an
 CC antibody; diagnosing a disease, disorder, syndrome, or condition
 CC comprising cancer, or proliferative, inflammatory, immune, metabolic,
 CC bone, CNS, genetic, bacterial and viral diseases, disorders, syndromes or
 CC conditions in a patient; a modulator composition comprising a modulator
 CC and a carrier; gene therapy; prophylactic or therapeutic treatment of a
 CC subject; an isolated modified cell comprising at least one first
 CC heterologous nucleic acid molecule, where the first heterologous nucleic
 CC acid molecule comprises a first polynucleotide sequence that encodes a
 CC first polypeptide; a non-human animal deficient in the polypeptide or
 CC that over-expresses the polypeptide; isolated tissues derived from the
 CC non-human animal; and one or more cells derived from the non-human
 CC animal. The nucleic acid is useful in preparing a composition for
 CC diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.
 CC This is the amino acid sequence of a novel human polypeptide of the
 CC invention.
 XX
 XX Sequence 216 AA;
 SQ
 Query Match 12.5%; Score 96; DB 8; Length 216;

Best Local Similarity 59.0%; Pred. No. -0.043;
Matches 23; Conservative 2; Mismatches 10; Indels 4; Gaps 1;

```

QY      114 MEWEVFPLEGRPEAGALQRLPQPSAL----LACRCAG 148
      ||||| ||||| ||| ||| ||||| :||| ||
Dp      1 MEWEGSPLESGRSAALAFLLPRPNSALFROSMACRHAG 39

```

RESULT 9
ABG23765
ID ABG23765 standard: protein: 87 AA:

XX
AC ABG23765;
XX
DT 18-FEB-2002 (first entry)

XX
DE
Novel human diagnostic protein #23756.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX

OS Homo sapiens.

XX PN WO200175067-A2.

XX
PD
11-OCT-2001XX
PF
30-MAR-2001:XX
PR 31-MAR-2000: 2000US-00540217.PR 23-AUG-2000; 2000US-00649167.
YY

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YF;
XX

DR WPI; 2001-639362/73.
DR N-PSDB: AAS87952.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 20: SEQ ID NO 54124; 103pp; English.
XX

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (I) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pub/published/pct/sequences

XX Sequence 87 AA;

Query Match 12.4%; Score 95.5; DB 4; Length 87;
Best Local Similarity 62.5%; Pred. No. 0.015;
Matches 20: Conservative 2; Mismatches 9; Indels

Best Local Similarity 62.5%; Pred. NO. 0.013;
Matches 20: Conservative 2: Mismatches 9: Indels 1: Gaps 1;

```

QY      110 ARRGMEWE-VFPLELGRPEAGALQRLPOPNSA 140
      ||:||||| ||||| ||||| ||||| |||||
Db      7 ARKGMEWEDLPLEFGHPTASLLSDRPOPNSS 38

```

RESULT 10

ABG12819
ID ABG12819 standard; protein; 87 AA.

AA ABG12819;

DT 18-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #12810.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

XX
QS
Homo sapiens.XX
PN
WO200175067-A2XX
11-OCT-2001XX
DE 30-MAR-2001.XX
BB 21-MAB-2000: 2000US-00540217

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX
PS Claim 20: SEO ID NO 43178; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at fto.wipo.int/pub/published pct sequences

XX Sequence 87 AA:

Query Match 12.1%; Score 93; DB 4; Length 87;

BEST LOCAL SIMILARITY	43.1%	FREQ. NO.	8,025
MATCHES	28;	CONSERVATIVE	
	5;	MISMATCHES	16;
		INDELS	16;
		GAPS	3;

90 VGLENERKV-LSGGSSOOMGEARRGMWVFFLELGR---PEAGALQRLPQNSALLACR 145

GenCore version 5.1.9
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2007, 14:57:46 ; Search time 299 Seconds

(without alignments)

464.055 Million cell updates/sec

Title: US-10-777-578-2

Perfect score: 770

Sequence: 1 MVVFTLLVLLLLLLLTALC.....LQRLPQNSALLACRCAGAY 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87.5	11.4	483	Q4HI99 CAMCO	Q4hi99 campylobact
2	85.5	11.1	550	Q8XR68_RALSO	Q8xt68 ralstonia s
3	83.5	10.8	254	Q97J11_CLOBAB	Q97j11 clostridium
4	82	10.6	404	Q9SELI_SOYBN	Q9seli glycine max
5	81	10.5	462	Q743Y7_MYCPA	Q743y7 mycobacteri
6	80.5	10.5	398	Q6AQH5_DESPS	Q6agh5 desulfotale
7	79.5	10.3	470	Q2NAE4_9SPHN	Q2nae4 erythrobact
8	79.5	10.3	953	Q82V63_NITEU	Q82v63 nitrosomona
9	79	10.3	414	Q4LU74_FERAC	Q4lu74 ferroplasma
10	78.5	10.2	673	Q8TFG0_SCHPO	Q8tfgo schizosacch
11	78	10.1	326	Q4VQ72_9REOV	Q4vq72 porcine rot
12	78	10.1	337	Q471X8_RALEJ	Q471x8 ralstonia e
13	78	10.1	337	Q894Q0_CLOTE	Q894q0 clostridium
14	77.5	10.1	305	Q8W2T3_ORYSA	Q8w2t3 oryza sativ
15	77.5	10.1	481	Q4WKL4_ASPPU	Q4wkl4 aspergillus
16	77	10.0	2005	Q5STG5_HUMAN	Q5stg5 homo sapien
17	76.5	9.9	956	Q5PB30_ANAMM	Q5pb30 anaplasmam
18	76.5	9.9	2604	Q50T90_ENTHI	Q50t90 entamoeba h
19	76.5	9.9	5120	PCLO_CHICK	Q9pu36 gallus gall
20	76	9.9	359	Q4J7S5_SULAC	Q4j7s5 sulfolobus
21	76	9.9	431	TOLB_MYXXA	Q84ff7 myxococcus
22	76	9.9	506	Q3QTN9_9RHOB	Q3qtn9 silicibacte
23	76	9.9	604	Q55PQ5_CRYNE	Q55pq5 cryptococcu
24	76	9.9	604	Q5KQD3_CRYNE	Q5kdq3 cryptococcu
25	76	9.9	1086	Q3PWT5_NITHA	Q3pwt5 nitrobacter
26	76	9.9	1097	Q9HCY0_HUMAN	Q9hcy0 homo sapien
27	75.5	9.8	507	CRNA_EMENI	P22152 emericella
28	75.5	9.8	507	Q5BEM2_EMENI	Q5bem2 aspergillus
29	75.5	9.8	550	Q976S2_SULTO	Q976s2 sulfolobus
30	75	9.7	116	Q5EY93_9EMBE	Q5ey93 ammodramus
31	75	9.7	400	Q9P8B4_AGABI	Q9p8b4 agaricus bi

32 75 9.7 405 2 Q4FNZ4_PELUB Q4fnz4 pelagibacte
33 75 9.7 497 2 Q5R5P1_PONPY Q5r5p1 pongo pygma
34 75 9.7 583 2 Q41VE2_DESHA Q41ve2 desulfitoba
35 74.5 9.7 319 2 Q4B7L2_BURVI Q4b7l2 burkholderi
36 74.5 9.7 423 2 Q32P40_HUMAN Q32p40 homo sapien
37 74.5 9.7 1153 2 Q7XUT7_ORYSA Q7xut7 oryza sativ
38 74 9.6 385 2 Q335T1_CROPO Q335t1 crocodylus
39 74 9.6 681 2 Q3TTV4_MOUSE Q3ttv4 mus musculus
40 74 9.6 1159 2 Q62120_ORYSA Q62120 oryza sativ
41 74 9.6 1188 2 Q69QL2_ORYSA Q69ql2 oryza sativ
42 74 9.6 1281 2 Q6TJY0_ORYSA Q6tjy0 oryza sativ
43 74 9.6 1302 2 Q69QM8_ORYSA Q69qm8 oryza sativ
44 74 9.6 1329 2 Q7RQ32_PLAYO Q7rq32 plasmodium
45 74 9.6 1479 2 Q3U306_MOUSE Q3u306 mus musculus

ALIGNMENTS

RESULT 1

Q4HI99 CAMCO PRELIMINARY; PRT; 483 AA.
AC Q4HI99;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Integral membrane protein MviN.
DE Name=mviN; ORFNames=CCO0860;
GN Campylobacter coli RM2228.
OS Campylobacter coli RM2228.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=306254;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RM2228;
RA Fouts D.E., Mongodin E.F., Mandrell R.E., Miller W.G., Rasko D.A.,
RA Jacques R.J., Brinkac L.M., DeBoy R.T., Parker C.T., Daugherty S.C.,
RA Dodson R.J., Durkin A.S., Madupu R.R., Sullivan S.A., Shetty J.U.,
RA Ayodeji M.A., Shvartsbeyn A.A., Schatz M.C., Badger J.H., Fraser C.M.,
RA Nelson K.E.;
RT "Major structural and novel potential virulence mechanisms from the
RT genomes of multiple Campylobacter species.";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; RAFL01000001; EAL57586.1; -: Genomic DNA.
CC GO; GO:0016021; C:integral to membrane; IEA.
CC GO; GO:0009405; P:patogenesis; IEA.
CC InterPro; IPR004268; MVIN_like.
CC Pfam; PF03023; MVIN; 1.
CC PRINTS; PR01806; VIRPACTRMVIN.
CC TIGRFAM; TIGR01695; mviN; 1.
CC SEQUENCE 483 AA; 54132 MW; 90C4513C4BBE71AB CRC64;

Query Match 11.4%; Score 87.5; DB 2; Length 483;

Best Local Similarity 33.3%; Pred. No. 4.7;

Matches 32; Conservative 13; Mismatches 38; Indels 13; Gaps 4;

QY 5 TLLVLLLLLLLTALCKAL-----SQSLPYTLRPSQSL--FLAITDKIDIQVFLP 56
DB 381 TAAFTAFKALLSALCSLVFLIKDESKVIAVALSSLSAFYLIIANIKERFGKNFFA 440

QY 57 LTGGKCLHLRLTGQAFCL--EFLPWCNGIIESLV 90

DB 441 LISFRCFLMIIGLVVFTLLIYEIKPY---LIEALL 473

RESULT 2


```

OS  Clostridium acetobutylicum.
OC  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC  Clostridium.
OX  NCBI_TaxID=1488;
RN  [1]
RP  NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX  STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX  MEDLINE=21359325; PubMed=11466286;
RX  DOI=10.1128/JB.183.16.4823-4838.2001;
RA  Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA  Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA  Tatuov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,
RA  Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RT  "Genome sequence and comparative analysis of the solvent-producing
RT  bacterium Clostridium acetobutylicum.";
RL  J. Bacteriol. 183:4823-4838(2001).
CC  - SIMILARITY: Belongs to the ABC transporter family.
CC  Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC  Distributed under the Creative Commons Attribution-NoDerivs License
CC  -----
CC  EMBL; AE007658; AAK79443.1; -; Genomic_DNA.
DR  PIR; H97081; H97081.
DR  HSSP; Q9YGA6; IG29.
DR  Biocyc; CACB1488:CAC1475-MONOMER; -.
DR  GO; GO:0016020; C:membrane; IEA.
DR  GO; GO:0005524; F:ATP binding; IEA.
DR  GO; GO:0016887; F:ATPase activity; IEA.
DR  GO; GO:0000166; F:nucleotide binding; IEA.
DR  GO; GO:0006810; P:transport; IEA.
DR  InterPro; IPR003593; AAA_ATPase.
DR  InterPro; IPR003439; ABC_transporter.
DR  InterPro; IPR001452; SH3_DOMAIN.
DR  Pfam; PF00005; ABC_tran; 1.
DR  PRINTS; PR00452; SH3DOMAIN.
DR  ProDom; PD000006; ABC_transporter; 1.
DR  SMART; SM00382; AAA; 1.
DR  PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR  PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW  ATP-binding; Complete proteome; Membrane; Nucleotide-binding;
KW  Transport.
SQ  SEQUENCE 254 AA; 28846 MW; 5172AEB84EEAD3BA CRC64;

Query Match 10.8%; Score 83.5; DB 2; Length 254;
Best Local Similarity 32.5%; Pred. No. 6;
Matches 3; Conservative 10; Mismatches 42; Indels 25; Gaps 6;

Qy 17 TALCKALSQLPYLYRPSQSLFLITDKKIDIQYFLPLTG-----GKCLHLRLTGQ 70
Db 45 TLLKMINR-----LYEPDKGSIVLFNEDIKKIDVKKLRRSIGVYIQVGLFPHMTIANN 99

Qy 71 RAFCILEFLPWCNGIIES-----LVGLE-NERK-----VLSGSSSQMGGEAR 111
Db 100 IA-TVPKLLKWDREIEKRIEDELHLVGLPEPNEPKRYPSQLSGGQQORIGLAR 152

RESULT 4
Q9SELI_SOYBN PRELIMINARY; PRT; 404 AA.
ID Q9SELI_SOYBN PRELIMINARY; PRT; 404 AA.
AC Q9SELI;
DT 01-MAY-2000, integrated into UniProtKB/TREMBL.
DT 01-MAY-2000, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Seed maturation protein PM23 (Fragment).
GN Name=PM23;
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
OC Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP NUCLEOTIDE SEQUENCE.

Q8XR68_RALSO PRELIMINARY; PRT; 550 AA.
AC Q8XR68;
DT 01-MAR-2002, integrated into UniProtKB/TREMBL.
DT 07-FEB-2006, sequence version 19.
DE PROBABLE CYTOCHROME C OXIDASE SUBUNIT TRANSMEMBRANE PROTEIN
DE (EC 1.9.3.11).
GN OrderedLocusNames=RS0993; ORNames=RS02325;
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cumac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AL646082; CAD18144.1; -; Genomic_DNA.
DR Biocyc; RSOL305:RSP0993-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0045278; C:respiratory chain complex IV (sensu Bacteria); IEA.
DR GO; GO:0004129; F:cytochrome-c oxidase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR004677; COXN.
DR InterPro; IPR000883; COX1.
DR PANTHER; PTHR10422:SF1; COXN; 1.
DR PANTHER; PTHR10422; COX1; 1.
DR Pfam; PF00115; COX1; 1.
DR PROSITE; PS00855; COX1; 1.
KW Complete proteome; Oxidoreductase; Plasmid; Transmembrane.
SQ SEQUENCE 550 AA; 58994 MW; 3D8A20B5F5F35C4C CRC64;

Query Match 11.1%; Score 85.5; DB 2; Length 550;
Best Local Similarity 28.6%; Pred. No. 8.6;
Matches 48; Conservative 19; Mismatches 62; Indels 39; Gaps 9;

Qy 1 MVVFTLL---LVLLLLLLLTKALQSPLPYLYRPSQSLFLITDKKIDIQY-FLP 56
Db 94 LVVFFVLCAMVWLLVASLAGITASLKLHLPDAL-ATQAWLSFGRIRTIHLNAVAYGAP 152

Qy 57 LTG-----GKCLHLRLTGORATCILEFLPWCNGIIESL-----VGLNERKVLGGSS 104
Db 153 MAGLGIAQFVLPRVLKTLVGGN-FAVLGAMMNVAGVGLSGIAGVLSL----- 201

Qy 105 QOMGEARGMEVPEFLGRPEA--GALQRLPQPNALLACRCAGAY 150
Db 202 -----GLEWLEIPWQIGVLFAVGGLVGIPLV-LTLATRVSHLY 240

RESULT 3
Q97J11_CLOAB PRELIMINARY; PRT; 254 AA.
ID Q97J11_CLOAB PRELIMINARY; PRT; 254 AA.
AC Q97J11;
DT 01-OCT-2001, integrated into UniProtKB/TREMBL.
DT 01-OCT-2001, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Proline/glycine betaine ABC transport system, ATPase component.
GN OrderedLocusNames=CAC1475;

```

```
RA Hsing Y.I.C., Lin T.Y., Lin T.Y., Liu S.M., Chow T.Y.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; AF116752; AAP21309.1; -; mRNA.  
FT NON TER  
SQ SEQUENCE 404 AA; 45221 MW; 684D285911DEF19E CRC64;  
  
Query Match 10.8%; Score 82; DB 2; Length 404;  
Best Local Similarity 23.0%; Pred. No. 14;  
Matches 41; Conservative 33; Mismatches 44; Indels 60; Gaps 8;  
  
QY 6 LLLVLLLLLLTALCKALSQSPY-----TLRPOSSLSFL----- 41  
Db 19 VLVVVLLIGMATLSLSSLPFLHRYNPNPNSLFRPSLSFSKTRSPFLVLAASS 78  
QY 42 -----LITDIKKIDIQVF--LPLTGGKCLHLRLTGQAFCLIFLFP--WC 82  
Db 79 HDPASNSKSVLTLEIQETPLDVSHIQKDVPTTADAMKRTISG-----MLGLPDSDF 133  
QY 83 NGIIESVLGLENKRVLSGGSSQOMGEARRGMW-----EVFLELGRPEAGALQ 132  
Db 134 HVVIEAL--WEPLSKLLI--SSMNTGYTLRNVEYRLCLEKNLDMFPGDIEKPKAESMK 187  
  
RESULT 5  
Q743Y7 MYCPA  
ID Q743Y7 MYCPA PRELIMINARY; PRT; 462 AA.  
AC Q743Y7;  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 11.  
DE Hypothetical protein.  
GN OrderedLocusNames=MAP0454; ORFNames=MAP_0454;  
OS Mycobacterium paratuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium avium complex (MAC).  
OX NCBI_TaxID=1770;  
RN (1)  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;  
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; AE016958; AAS0271.1; -; Genomic_DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 462 AA; 50122 MW; 437D61E14569F7AD CRC64;  
  
Query Match 10.5%; Score 81; DB 2; Length 462;  
Best Local Similarity 25.2%; Pred. No. 21;  
Matches 38; Conservative 25; Mismatches 52; Indels 36; Gaps 6;  
  
QY 2 VVFTLLVLLLLLALCKALSQSLPYTLRPSQSLFLITDIKKI-DIQVFLPTGCG 60  
Db 36 LVFTNRVELLKLAVILAAVAGAFVSVLYRQSDAQSRVRLKLVYDLQ----- 87  
QY 61 KCLHLRLTGQAFCLIFLWPNCNGIIESVLGLENKRVLSGGSSQ-----QMGEA 110  
Db 88 --LDREISARREY-----ELTVESQRLRELASELRAQAADDLAELRAELSAL 132  
QY 111 RGMWEVFPLELG-RPEAGALQRLPQNSA 140  
Db 133 RTSLE-ILFDTDLGQALGAFEGEPQPERA 162  
  
RESULT 6  
Q6AQH5_DESPS
```

```
ID Q6AQH5_DESPS PRELIMINARY; PRT; 398 AA.  
AC Q6AQH5;  
DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.  
DT 13-SEP-2004, sequence version 1.  
DT 07-FEB-2006, entry version 8.  
DE Probable Na+/H+ antiporter.  
GN OrderedLocusNames=DP0669;  
OS Desulfotalea psychrophila.  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacteriales;  
OC Desulfobacteraceae; Desulfotalea.  
OX NCBI_TaxID=84980;  
RN (1)  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=LSV54 / DSM 12343;  
RX PubMed=15305914; DOI=10.1111/j.1462-2920.2004.00665.x;  
RA Rabus R., Ruepp A., Frickey T., Rattei T., Fartmann B., Stark M.,  
RA Bauer M., Zibat A., Lombardot T., Becker I., Amann J., Gellner K.,  
RA Teeling H., Leuschner W.D., Gloeckner F.-O., Lupas A.N., Amann R.,  
RA Klenk H.-P.;  
RT "The genome of Desulfotalea psychrophila, a sulfate-reducing bacterium  
RT from permanently cold Arctic sediments.";  
RL Environ. Microbiol. 6:887-902(2004).  
CC  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; CR522870; CAG35398.1; -; Genomic_DNA.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0006885; P:regulation of pH; IEA.  
DR GO; GO:0006814; P:sodium ion transport; IEA.  
DR InterPro; IPR004670; NhaA.  
DR Pfam; PF06965; Na_H.antiport.1; 1.  
DR TIGRFAMs; TIGR00773; NhaA; 1.  
KW Complete proteome.  
SQ SEQUENCE 398 AA; 43064 MW; 7571BD49177DC42E CRC64;  
  
Query Match 10.5%; Score 80.5; DB 2; Length 398;  
Best Local Similarity 28.2%; Pred. No. 20;  
Matches 42; Conservative 18; Mismatches 42; Indels 47; Gaps 7;  
  
QY 3 VFTLL-----LVLLLLLLTALCKALSQSLPYTLRPSQSLFLITDIKKIDIQVFLPT 58  
Db 10 VFALLKSNISGGILLMLATALLIMANSPGHVLY-----SMLITTPVE---VRFGLPE 59  
QY 59 GSKCLHLRLTGQAFCLIFLWPNCN-GIIES---LVGLENKRVLSGGSSQOMGEARRGM 114  
Db 60 TAKPL-----LLWINDGLMAGFFLVGLKREIFEGGLSQRS----- 97  
QY 115 EWEVFPLELGRPEAGALQRLPQNSALLA 143  
Db 98 -----NILLPAIGALGMVVPSCIYLA 119  
  
RESULT 7  
Q2NAE4_9SPHN  
ID Q2NAE4_9SPHN PRELIMINARY; PRT; 470 AA.  
AC Q2NAE4;  
DT 07-FEB-2006, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, sequence version 1.  
DT 07-FEB-2006, entry version 1.  
DE NtrC, nitrogen regulation protein NtrC.  
GN ORFNames=ELI_06275;  
OS Erythrobacter litoralis HTCC2594.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;  
OC Erythrobacteraceae; Erythrobacter.  
OX NCBI_TaxID=314225;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=HTCC2594;  
RA Giannoni S.J., Cho J.-C., Ferreira S., Johnson J., Kravits S.,  
RA Halpern A., Remington K., Beeson K., Tran B., Rogers Y.-H.,  
RA Friedman R., Venter J.C.;  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
```

```
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000157; ABCG3347.1; -; Genomic DNA.
SQ SEQUENCE 470 AA; 51111 MW; 7DC67C079538405F CRC64;

Query Match 10.3%; Score 79.5; DB 2; Length 470;
Best Local Similarity 27.7%; Pred. No. 30;
Matches 28; Conservative 13; Mismatches 37; Indels 23; Gaps 4;

Qy 52 QYFLPLTGGKCLHLRTGQAFCLFELPWCNGIIE-----SLVGLNERK 97
Db 320 RHFLALADEGLPLRTIDEAIAVLROQPWKGNVRLNLYRLALLARNVDVDAELAE 379

Qy 98 VL--SGSSQOMGEA---RRMEWEVFPLEGRPEAGALOR 133
Db 380 VLDHEGGPSTMEGPPSSVDRLDW-----IENRRPAPGALYR 416

RESULT 8
Q82V63 NITEU PRELIMINARY; PRT; 953 AA.
AC O82V63;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Putative oxygenase.
GN OrderedLocusNames=NE1238;
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 19718 / IF0 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA DOI=10.1128/JB.185.9.2759-2773.2003;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.L.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RA "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BX321850; CAD85149.1; -; Genomic DNA.
DR Biocyc; NEUR228410:NE1238-MONOMER; -.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR002007; Anim_peroxidase.
DR Pfam; PF03098; An_peroxidase; 1.
DR PROSITE; PS0292; PEROXIDASE_3; 1.
KW Complete proteome.
SQ SEQUENCE 953 AA; 108546 MW; B2461487B7E3EA66 CRC64;

Query Match 10.3%; Score 79.5; DB 2; Length 953;
Best Local Similarity 31.2%; Pred. No. 65;
Matches 44; Conservative 16; Mismatches 38; Indels 43; Gaps 9;

Qy 4 FTLLVLLLLLLLLLTCALKSQSLPYTLRPOSSLSFLITDIK---KIDIOYFLPTGG 60
Db 4 FPIVIVLGLLLSG-CDALEPEI-----SCLSVLMQGDIOHVAKTREQRFJ-----G 50

Qy 61 KCLHLRLTGQAFCL-----LEFLPCN-----GIIESLVGLNERKVLGG--SSQ 106
Db 51 K-----VTGRRACGGGDAVALNRPWLDFNFWGTGDSL-----SLSSSLASSF 97

Qy 107 MGEARRGMWEVFPLEGRPE 127
Db 98 FGPNERGINSALYELQRIE 118
```

```
RESULT 9
Q41U74 FERAC PRELIMINARY; PRT; 414 AA.
AC Q41U74;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=FacidRAFT_1367;
OS Ferroplasma acidarmanus Ferl.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Ferroplasmaceae; Ferroplasma.
OX NCBI_TaxID=333146;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=Ferl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Ferroplasma
RT acidarmanus Ferl.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=Ferl;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Ferroplasma acidarmanus
RT Ferl.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=Ferl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAC04000003; EAM94257.1; -; Genomic DNA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001046; Nramp.
DR ProDom; PD001861; Nramp; 1.
DR Hypothetical protein.
KW SEQUENCE 414 AA; 44916 MW; 2A80611B1E404133 CRC64;

Query Match 10.3%; Score 79; DB 2; Length 414;
Best Local Similarity 22.3%; Pred. No. 30;
Matches 33; Conservative 30; Mismatches 57; Indels 28; Gaps 5;

Qy 1 MWFTLLVLLLLLLLLLTCALKSQSLPYT---LYRPOSSLSFLITDIKIDIOYFL-- 55
Db 153 LIIISLLLIILLLSL-ALRGIIPVSSPLANPILIKTAGYFFLLAANVGAVIMPFMIF 211

Qy 56 -----PLTGKCLHLRTGQAFCLFELPWCNGIIESLV-----GLENER 96
Db 212 QASATGKLGIDGG---HIRITRRSLRIRKETLAGAIVTELLMWIAEMAFAGIPHAS 268

Qy 97 KVLGGSSQQMGGEARRGMWEVFPLELG 124
Db 269 HSSFFATPQQLGKVLIPVAGDFSPILLG 296

RESULT 10
Q8TFGO_SCHPO
```

```

ID Q8TFGO SCHPO PRELIMINARY; PRT; 673 AA.
AC Q8TFGO;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE SPBP886.02c protein (Spap886.02c protein) (spap886.02c protein).
GN ORFNames=SPAP886.02c, SPAP886.02c, SPBP886.02c;
OS Schizosaccharomyces pombe (Fission yeast)
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=972;
RX MEDLINE=21948401; PubMed=11859360; DOI=10.1038/nature724;
RA Wood V., Guilliam R., Rajandream M.A., Lyne M.H., Lyne R., Stewart A.,
RA Sgouros J.C., Peat N., Hayles J., Baker S.G., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D.E., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jegels K.,
RA James K.D., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K.L., Murphy L.D., Niblett D., Odeil C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K.M., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M.N., Squares R., Squares K., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J.R., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Duesterhoef A., Fritz C., Holzer E., Moestl D.,
RA Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R.,
RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Porashkin J.,
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AL691490; CAD27908.1; -; Genomic_DNA.
DR GeneDB Spombe; SPBP886.02c; -;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001734; Na/solut_sympor.
DR PANTHER; PTHR11819; Na/solut_sympor; 1.
DR PROSITE; PS0283; NA_SOLUT_SYMP_3; 1.
KW Complete proteome.
SQ SEQUENCE 673 AA; 72604 MW; 406931E32F9F0041 CRC64;

Query Match 10.2%; Score 78.5; DB 2; Length 673;
Best Local Similarity 22.9%; Pred. No. 56;
Matches 33; Conservative 24; Mismatches 52; Indels 35; Gaps 7;

OY 6 LLLVLLVLLVLTALCKALQSLSPTLYRPOSSLSFLITDIKKIDIQFLPTGGKCLHL 65
DB 197 MYIILVTCYTVYSSSLIGSPAKY-----DMUK-EVQVYPATGQS-YL 242
OY 66 RLTGQAFCLFLFPW--CNGIIESLVGLE--NKKVLSGGSSQMGARRGMEVFPPL 121
DB 243 SFKNSE----WNYLWTSWIMGLSSVFGDPGYSQRAIASDAKSVFGYLMGLCWIIIPM 298
OY 122 ELGRPEAGALQLPQNSALLACR 145
DB 299 ALG-----SSAGLACR 309

RESULT 11
OAVQ72_9REOV PRELIMINARY; PRT; 326 AA.
ID OAVQ72_9REOV
AC OAVQ72;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Outer capsid protein VP7.
DS Porcine rotavirus.
OS Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10913;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=34461-4;
RX PubMed=15914225; DOI=10.1016/j.virol.2005.03.031;
RA Martella V., Ciarlet M., Baselga R., Arista S., Elia G., Lorusso E.,
RA Banyai K., Terio V., Madio A., Ruggeri F.M., Falcone E., Camero M.,
RA Decaro N., Buonavoglia C.;
RT "Sequence analysis of the VP7 and VP4 genes identifies a novel VP7
RT gene allele of porcine rotaviruses, sharing a common evolutionary
RT origin with human G2 rotaviruses."
RL Virology 337:111-123(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AY766085; AAX31515.1; -; Genomic_RNA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR InterPro; IPR001963; VP7.
DR Pfam; PF00434; VP7; 1.
DR ProDom; PD000191; VP7; 1.
DR PROSITE; PS00434; VP7; 1.
SQ SEQUENCE 326 AA; 36993 MW; E44FAF13B634DD9D CRC64;

Query Match 10.1%; Score 78; DB 2; Length 326;
Best Local Similarity 22.6%; Pred. No. 29;
Matches 24; Conservative 24; Mismatches 32; Indels 26; Gaps 5;

OY 5 TLLVLLVLLVLTALCKALQSLSPTLYRPOSSLSFLITDIKKIDIQFLPTGG 59
DB 8 TILTLISIIILNVLTKVTSTMDYIYR-----FLLLIALISPLVKSQNGVILPTG 61
OY 60 G-KCLHLRLTQRAF-----CI-----LEFLPWCNGIIESLV 90
DB 62 SMDAVYANSTSEKFLTLTCIYVYPTAKNEISDTWNTLSQLLL 107

RESULT 12
O471X8_RALEJ PRELIMINARY; PRT; 337 AA.
ID O471X8_RALEJ
AC O471X8;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE DNA polymerase III delta prime subunit (EC 2.7.7.7).
GN OrderedLocustNames=Reut Al435;
OS Ralstonia eutropha (strain JMPL134) (Alcaligenes eutrophus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Cupriavidus.
OX NCBI_TaxID=264198;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Irani S., Pitluck S., Goltsman E., Martinez M.,
RA Schmutz J., Larimer F., Land M., Lykidis A., Richardson P.;
RT "Complete sequence of chromosome 1 of Ralstonia eutropha JMPL134."
RL submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; CP000090; AAZ60805.1; -; Genomic_DNA.
DR GO; GO:0043234; C:protein complex; IEA.
```

DR GO: GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003987; F:DNA-directed DNA polymerase activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006260; P:DNA replication; IEA.
DR InterPro: IPR004622; HolB.
DR InterPro: IPR000862; REC.
DR TIGRFAMs: TIGR00678; holB; 1.
KW Complete proteome; Nucleotidyltransferase; Transferase.
SQ SEQUENCE 337 AA; 36953 MW; 96D553A3655145F5 CRC64;

Query Match 10.1%; Score 78; DB 2; Length 337;
Best Local Similarity 26.8%; Pred. No. 30;
Matches 51; Conservative 18; Mismatches 55; Indels 66; Gaps 10

QY 8 LVLLLLLLTALCKALQSPLYTYLRPOSSLSFLITDIKKIDIQYFLPTGGKCLHL-- 65
DB 124 LRVVVVPLDALQVEGANALLKTLLEEPSSTVFLLVTD---RLDRILPTILSRCRQFST 179
QY 66 -RLTGQAFCLILEFLPW--CNGIETSLVGLNERKVLSSGGSQMGEARRGMEWVFPLE 122
DB 180 QRPTPEAA-----LAWLKSGQVADAEAAQL-----ALAGGSP-----TALHAAEAEEQPLQ 225
QY 123 ---LGRPEAGA-----LORLPQP-----NSALLAC----- 144
DB 226 RNLVQGLGAGAEFATGAEEQLKLPVAVILQIRWTYDLLKRLDASGNVAPRYFPKE 285
QY 145 -----RCAGA 149
DB 286 RTVLARCAGA 295

RESULT 13
Q894Q0_CLOTE
ID ID Q894Q0_CLOTE PRELIMINARY; PRT; 392 AA.
AC Q894Q0;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Na+ ABC transporter natB.
GN Name=natB; OrderedLocusNames=CTC01485; ORFNames=CTC_01485;
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;
RA Brueggemann H., Baeumer S., Fricke W.F., Wierer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gertschak G.
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease."
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EBI; AEO15927; AAO36042.1; -; Genomic_DNA.
DR DR BiOCyc: CTET212717:CTC01485-MONOMER; -.
KW Complete proteome.
SQ SEQUENCE 392 AA; 43559 MW; 7DB3F993979FBBDE CRC64;

Query Match 10.1%; Score 78; DB 2; Length 392;
Best Local Similarity 31.0%; Pred. No. 35;
Matches 31; Conservative 15; Mismatches 26; Indels 28; Gaps 5

QY 5 TLLVLLVLLLLLTALCKALQSLSL-----PYTYLRPOSSLSFLLI-----TDIKKID 50
DB 274 TLVLVNLIPILTYVFGALQALSIYAKSFKAQTYLSITVSVMLIIYVWMMKDPKNIG 333
QY 51 IQYF-LPLTGGKCL-----HLRLT-GORAFCIL 76

Search completed: January 10, 2007, 15:06:02
Job time : 303 secs

FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/11/434,127
CURRENT FILING DATE: 2006-05-16